DOCUMENT RESUME

ED 185 976 IR 008 301

AUTHOR Long, Philip L.
TITLE Study of Message Text Formats: Bibliographic Search Queries.
INSTITUTION Library of Congress, Washington, D.C. Network Development Office.
PUB DATE 79
NOTE 42p.; For related documents, see ED 168 515-518. Network Planning Paper 5.
AVAILABLE FROM Customer Services Section, Cataloging Distribution Service, Library of Congress, Navy Yard Annex, Building 159, Washington, DC 20541

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Cataloging; *Information Networks; *Information Retrieval; Information Services; Information Systems; Library Networks; *Online Systems; *Search Strategies

ABSTRACT

This report presents an analysis of the search query arguments of several major bibliographic utilities and shows the commonalities and differences of the query structure among the various systems. The report illustrates a symbolic metalanguage that was developed to describe and compare the search arguments. In addition, status and error messages used in conjunction with search queries that should be incorporated in the National Commission on Libraries and Information Science (NCLIS) National Bureau of Standards (NBS) Applications Level Protocol are specified. The report recommends that certain search argument formats be standardized for computer-to-computer communications and that more work be performed in the analysis of bibliographic response formats on the various systems. Data about the following systems are included: the Library of Congress, the National Library of Medicine, OCLC, the University of Chicago, the Washington Library Network, Northwestern University, and the Research Libraries Information Network. (Author/FM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED185976

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ISSN 0160-9742

NETWORK PLANNING PAPER

NETWORK DEVELOPMENT OFFICE

NUMBER 5

DATE 1979

TITLE Study of Message Text Formats: Bibliographic Search Queries

Prepared by Philip L. Long

Edited and Revised by David C. Hartmann

Library of Congress

Washington

IR008301

PREFACE

This report presents an analysis of the search query arguments of several major bibliographic utilities and shows the commonalities and differences of the query structures among the various systems. The report presents a symbolic metalanguage that was developed to describe and compare the search arguments. In addition, status and error messages used in conjunction with search queries that should be incorporated in the NCLIS/NBS Applications Level Protocol are specified. The report recommends that certain search argument formats be standardized for computer-to-computer communications and that more work be performed in the analysis of bibliographic response formats on the various systems.

The document is based upon the final report of a study conducted by Philip L. Long of Philip Long Associates, Inc. The study was commissioned by the Committee for the Coordination of National Bibliographic Control (CCNBC) in response to a request from the Network Technical Architecture Group (NTAG) to have a study performed to assist in the initial planning of the emerging national library and information service network. CCNBC is involved in investigating some of the common problems for the information community at large and is financially supported by the National Science Foundation, the National Commission on Libraries and Information Science, and the Council on Library Resources. Opinions, findings, conclusions, or recommendations expressed herein do not necessarily reflect the views of the funding agencies.

David C. Hartmann, Network Development Office, Library of Congress, edited and revised the report and the members of NTAG updated the initial work to reflect system changes at their institutions. The search arguments presented in Section 4.0 of the report include formats at the institutions as of November 1978. The changes that have occurred since that time are not reflected in this study.

I would like to express my thanks to the members of NTAG who assisted in reviewing and updating the results of this study and to the staff members of the institutions that provided information for this study. I would especially like to thank Mary Jane Reed of the Washington Library Network, who provided assistance in the preparation of the search format matrix.

Henriette D. Avram
Director, Network Development Office
Library of Congress

October 1979

CONTENTS

Preface iii

1.0 Background 1

1.1 The Emerging National Network
1.2 Applications Level Protocol
1.3 Message Text Study

2.0 Specification of Search Catagories 6

3.0 Search Argument Metalanguage 7

4.0 Search Argument Formats 9

5.0 Status/Error Message Format 21

5.1 Status Messages
5.2 Error Messages
5.3 Suspend Messages
5.4 Informatory Messages

6.0 Search Query Responses 24

6.1 Response Contents
6.2 Response Formats

7.0 Recommendations and Conclusions 26

7.1 Standard Queries
7.2 Response Formats
7.3 Conclusions

1.0 Background

Since the early 1960s on-line computer services have been under development to answer the bibliographic needs of the library and information service community. These services emerged independently to support a wide variety of users--multi-institutions, single institutions, and in some cases specific divisions within a single institution. The services support specific needs such as cataloging, reference, serials control, and circulation, and because these systems were developed independently to support different functions the techniques employed were diverse. Various hardware configurations and software packages were used or developed for different reasons: optimally meeting the user's requirements, being most cost effective, using available systems and equipment, reducing development time, minimizing support and programming staff, or being compatible with other systems. For whatever reasons, a widely diverse set of computerized bibliographic services currently exist.

Almost all the systems do internal processing in different ways. A system which is reference oriented may sacrifice the capability of on-line update to allow for rapid access to information via efficient search and retrieval methods. Internal storage techniques differ from system to system, so that some systems duplicate certain data to optimize retrieval and display time while others use complex file design to maintain data integrity or allow easy file maintenance. Users of different on-line systems are most acutely aware of the differences in user interface techniques. Systems performing the same or similar functions have different command and response syntax and semantics. Error and control messages have different formats. Users of certain systems require extensive training to efficiently use the systems capabilities, and users of more than one system must be trained for each one of them, learning the various nuances of the command/response languages. As an individual uses these different systems, he must perform a mental conversion from system to system. Because systems provide different services, we see more and more people having to use multiple systems in performing their work and thus being forced to learn and use the multiple command/response languages.

1.1 The Emerging National Network

During the past several years an effort has been under way to develop a National Library and Information Service Network. In its report Toward a National Program for Library and Information Services: Goals for Action, the National Commission on Libraries and Information Science (NCLIS) recommended a national program, one of whose objectives was to "plan, develop and implement a nationwide network of library and information service."1/ The Network Advisory Committee (NAC), a committee of senior representatives from ongoing systems, was established at the Library of Congress to explore the requirements and

possibilities for increased cooperation among the components of the evolving national system. During its early meetings the NAC prepared a paper entitled Toward a National Library and Information Network: The Library Bibliographic Component that described some initial considerations for the library bibliographic component of the network. In this paper it identified a task to "design the technical network configuration or architecture" and recommended the formation of a small task force of technical staff from library networking organizations with operational automated systems. Such a task force was established in January 1977 and was called the Network Technical Architecture Group (NTAG). The group investigated some of the technical issues associated with networking and developed projects leading toward the creation of a national network. In addition, NTAG performed studies and prepared working papers investigating network architecture, network front end processors, value added networks, network technical design criteria, and networking standards.

1.2 Applications Level Protocol

In parallel to the work of NTAG, NCLIS, with support from the National Bureau of Standards (NBS), brought together a task force which developed a computer protocol for library and information services. The NCLIS/NBS Protocol is a computer-to-computer applications level protocol which allows for consistent and reliable communications between bibliographic applications operating in different computers. The structure and procedures for establishing sessions, exchanging messages, and controlling interactions between host computer systems are described in the document A Computer Network Protocol for Library and Information Applications. The protocol creates a vehicle by which sessions can be established so that search queries and responses can be exchanged, interlibrary loan requests sent, cataloging updates transmitted, and other library applications supported. The protocol does not, however, define the format of these "data type" messages but rather creates a mechansim by which they can be transmitted.

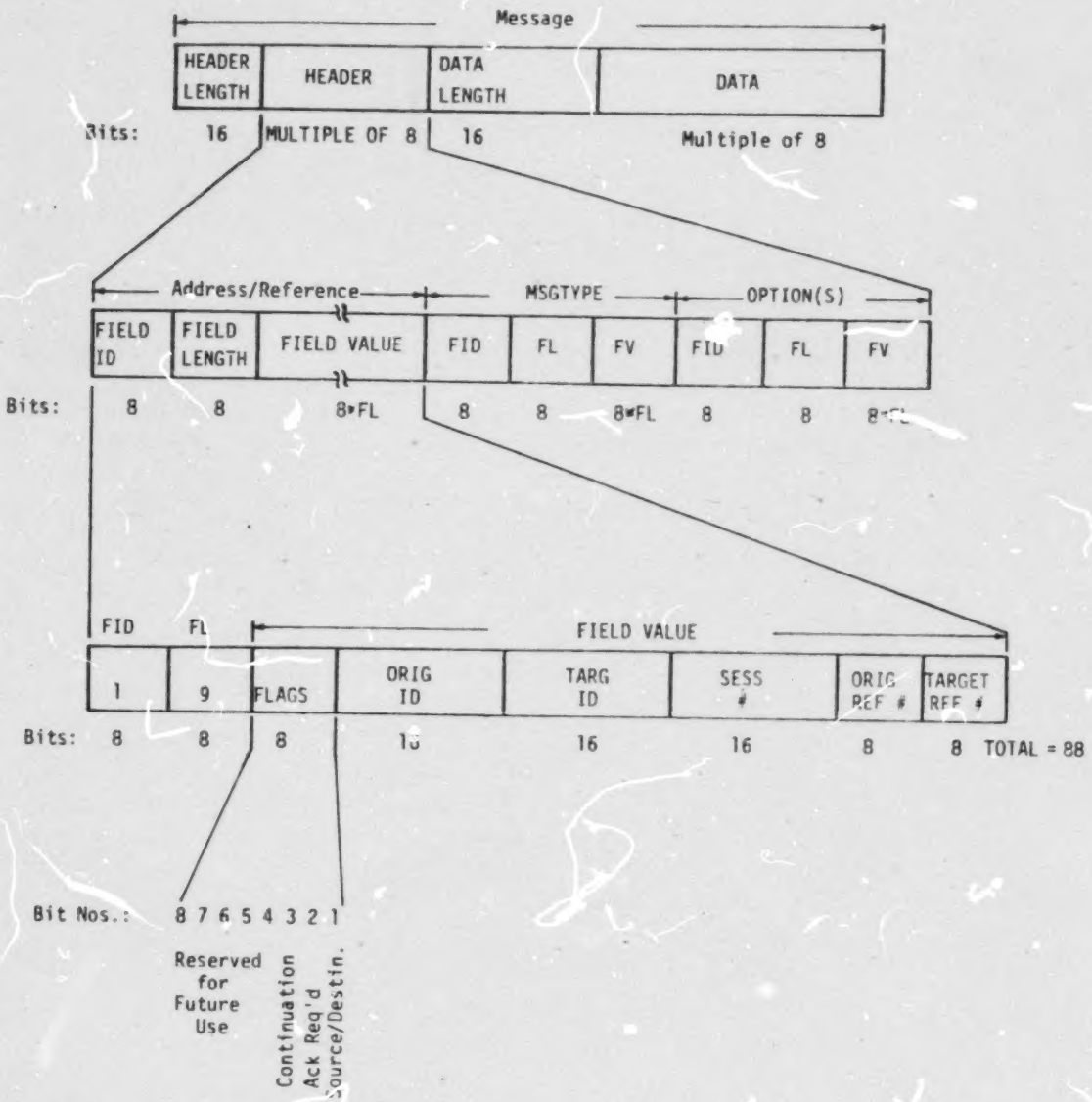
Figure 1.1 shows a message that is divided into its header portion and data portion, the header being subdivided into header fields. By specifying values in header fields, a host computer gives information so that another host can either process a message or realize it cannot handle the message and respond with an error message.

The NCLIS/NBS Task Force and NTAG both recognized that while the protocol as written answers some of the needs for exchanging messages in a nationwide network, it does not address the area of standard command formats and structures such as standard search queries, error messages, and cataloging commands.

> "Most likely, host processes will have differing commands. Therefore, some translating or reformatting of requests will be necessary. Translation could occur in any one of three ways:

Figure 1.1

Detailed Message Structure



SOURCE: NCLIS/NBS Task Force on Computer Network Protocol, A Computer Network Protocol for Library and Information Science Applications (Washington, 1977), p. 24.

1) The end user at the terminal could re-input the request in the format required by the remote host; 2) the local host automatically could translate the request into the required formats; 3) the local host automatically could translate the request into a Network standard format, which would in turn be translated by the remote host into its internal format. In the long run the third approach is preferable, but in the absence of necessary command standards, the application protocol will allow all three methods to be implemented." 2/

"Standardization of message content will be an important issue for the useful implementation of the protocol on a nationwide basis. Such standardization will also be very useful for initiating the Registry. The Task Force recommends that further work be done in this area." 3/

1.3 Message Text Study

During its regular meetings, NTAG, like the NCLIS/NBS Task Force, felt that some initial work should be done in the area of message text formats and asked the Committee for the Coordination of National Bibliographic Control to help fund an initial study. The objectives of this study were to survey existing bibliographic systems, collect applications message query and response formats for appropriate catagorization and classification, compare the formats and note any commonality in the hope of developing a set of standard network commands. Information was gathered about the systems at the following institutions:

Library of Congress

National Library of Medicine

OCLC, Inc.

University of Chicago

Washington Library Network

Northwestern University

BALLOTS (now Research Libraries Information Network)

User manuals and system documentation for the various systems were reviewed and user support personnel at each site were contacted to get a clear understanding of how each system operated and how the various messages were formatted. The messages were then analyzed and a metalanguage to describe the formats was developed. The following sections describe the results of this work.

NOTES:

1. National Commission on Libraries and Information Science, Toward a National Program for Library and Information Services: Goals for Action (Washington, 1975), p. 48.

2. NCLIS/NBS Task Force on Computer Network Protocol, A Computer Network Protocol for Library and Information Science Applications (Washington, 1977), p. 14.

3. Ibid., p. 47.

2.0 Specification of Search Categories

Computerized bibliographic systems have one or more fields within the bibliographic record that are searchable by a user of the system. These searchable fields, called access points, are usually based on the standard access points used in most card catalogs, although computerization has allowed the creation and sophisticated manipulation of more access points. The most common access points for the various systems were selected for comparison of the search terms used, and the format of search arguments used in the various systems was investigated. The format of the search argument for an access point was analyzed, rather than the access point's relationship to the work (that is, the formats of the search arguments for personal names used as access points in the various systems were compared, regardless of whether the personal name was the author or subject). Systems vary as to which fields can be used as access points--author (personal, corporate, and conference), series (title, corporate, and conference) and subject (topical, geographic, personal, corporate, and conference); however, within each system the search argument format of the type of access point remains fairly consistent.

The types of access points investigated were:

Personal Name (author or subject)

Corporate/Conference Name (author, series or subject)

Title (including series)

Author/Title

Subject Heading (topical and geographic)

Library of Congress Card Number (LCCN)

International Standard Book Number (ISBN)

International Standard Serial Number (ISSN)

Coden

Call Number

3.0 Search Argument Metalanguage

In order to compare the search arguments used on the various sytems, a metalanguage was developed to describe the format and syntax of the search arguments. The representation used to describe search arguments is as follows:

〈〉 indicates a string

[ ] indicates optional data

w indicates a word

c indicates an alphanumeric character

a indicates an alphabetic character

n indicates a numeric character

# indicates implicit or explicit truncation

Superscripts indicate (1) mandatory length when they occur alone
(2) maximum length when they occur with a subscript (when the exact value of superscripts varies with the data in the search argument, a letter value (m,n,i,j,k) is used)

Subscripts indicate minimum length.

Example: $[a_1^3]n^2\text{-}n_1^6$

represents a string with (optionally) one to three leading alphabetic characters, followed by exactly two numbers, followed by a hyphen, followed by one to six numbers. This is a typical format for an LC Card Number search. (ph78-12345)

Systems sometimes combine the search argument with a system command. (FIND, SHOW AUTH=, etc.) This study did not investigate the commands used on the various systems but rather assumed that as the NCLIS/NBS protocol was implemented, a series of message type codes and data format codes would be registered indicating the different types of searches. The message type field within the header portion of the message allows the assigning of a unique number to the different types of messages (search, update, etc.) and the data format field indicates the format of the data in the data portion of the message ("3,3 key", LCCN, etc.) The two header fields used in conjunction with one another would identify the type of search desired. As the protocol is

implemented the specific values indicating the types of message, data formats, etc., will be registered so that each system would understand the meaning of the message. The concept of a "Registry" of authorized values is described in the NCLIS/NBS Protocol document.

There are several approaches to assigning values to message type and data format fields. In one approach a unique numeric code in the message type field could be assigned for a "Bibliographic Search" type message. Several numeric codes in the data format field could be assigned for the types of search query. (Author-Personal Name-Keyword, Author-Corporate/Conference-Key, Title-Direct Order, Title-Keyword, Title Key, etc.) In another approach different codes in the message type field would be assigned for different access points (Author, Personal, Corporate/Conference, Title, Subject, etc.) and in the data format for different formats of searches (Keyed, Keyword, Direct Order, etc.). The chosen approach should be coordinated with the various systems.

4.0 Search Argument Formats

In this section the formats of the search arguments on the various systems are presented. The formats are described using the metalanguage defined earlier. To assist the reader in understanding the arguments, three sample bibliographic records are used to present examples of the search arguments. Figures 4.1 - 4.3 show the sample records. Figures 4.4 - 4.11 show the search argument formats for the eight different systems with sample arguments taken from examples 1, 2, and 3. Figure 4.12 summarizes the search arguments for the eight systems.

Example 1

Johnson, Ron, 1937-
The early sculpture of Picasso, 1901-1914 / Ron Johnson --
New York : Garland Pub., 1976.
iv, 245 p. : ill. ; 22 cm. -- (Outstanding dissertations in the fine arts)
Originally presented as the author's thesis, University of California, 1971.
Bibliography: p. 181-187.
ISBN 0-8240-1990-3
1. Picasso, Pablo, 1881-1973. I. Title. II. Series.
NB553.P45J63 1976 730'.92'4 75-23795
MARC

```
010    75-23795
020    0824019903
050 0  NB553.P45 $b J63 1976
082    730/.92/4
100 10 Johnson, Ron, $d 1937-
245 14 The early sculpture of Picasso, 1901-1914 / $c Ron Johnson.
260 0  New York : $b Garland Pub., $c 1976.
300    iv, 245 p. : $b ill. ; $c 22 cm.
440  0 Outstanding dissertations in the fine arts
500    Originally presented as the author's thesis, University
       of California, 1971.
504    Bibliography: p. 181-187.
600 10 Picasso, Pablo, $d 1881-1973.
```

Figure 4.1

Example 2

Ellis, John Tracy, 1905-
American Catholicism. [Chicago] University of Chicago Press [1956]
xiii, 207 p. 21 cm. (The Chicago history of American civilization) (Charles R. Walgreen Foundation. Lectures)
Bibliographical references included in "Notes" (p. 160-180)
"Suggested reading": p. 188-197.
1. Catholic Church in the United States--History.
I. Title. II. Series. III. Series: Chicago. University. Charles R. Walgreen Foundation for the Study of American Institutions. Lectures.
BX1406.E4 282.73 56-11002
MARC

```
010    56-11002
050 0  BX1406 $b .E4
082    282.73
100 10 Ellis, John Tracy, $d 1905-
245 10 American Catholicism.
260 0  [Chicago] $b University of Chicago Press $c [1956]
300    xiii, 207 p. $c 21 cm.
440  4 The Chicago history of American civilization
490 1  Charles R. Walgreen Foundation. Lectures
504    Bibliographical references included in "Notes" (p. 160-180)
       "Suggested reading": p. 188-197.
610 20 Catholic Church in the United States $x History.
810 1  Chicago. $b University. $b Charles R. Walgreen Foundation
       for the Study of American Institutions. $t Lectures.
```

Figure 4.2

Example 3

American Society for Information Science.
Journal of the American Society for Information Science
v. 21- Jan./Feb. 1970- [Washington, ASIS]
v. ill. 28 cm. bimonthly
Continues American Documentation.
ISSN 0002-8231
1. Information science--Periodicals. 2. Documentation--Periodicals.
Z100[illegible]477 020'.5 75-640174
MARC-S

```
C10    75-640[illegible]
022    0002-8231
030    AISJB6
050 0  Z1007 $b .A477
082    020/.5
110 20 American Society for Information Science.
222 00 Journal of the American Society for Information Science
245 00 Journal of the American Society for Information Science.
260 00 [Washington, $b ASIS]
300     $b ill. $c 28 cm.
362 0  v. 21-    Jan./Feb. 1970-
580    Continues American documentation.
650  0 Information science $x Periodicals.
650  0 Documentation $x Periodicals.
780 10  $t American documentation; a quarterly review of ideas,
        techniques, problems and achievements in documentation
        $x 0096-946X
```

Figure 4.3

BALLOTS

Access Point	Search Key	Example 1	Example 2	Example 3
Personal Name (Author, Subject)	$([c_1^n[\#]])_1^m\ c_3^n[\#]$ or $([w[\#]])_1^m\ c_3^n[\#]$, $([c_1^n[\#]])_1^m$	Ron Johnson Johnson, Ron Picasso, Pablo	J T Ellis Ellis, John T	
Corporate/Conference Name (Author,Series,Subject)	$<(w[\#]_)_1^n>$		Chicago University	American Society Information
Title	$<(w[\#])_1^n>$	Sculpture Picasso	American Catholicism	Journal Information Science
Author/Title	Boolean Coordination			
Series Title	$<(w[\#])_1^n>$	outstanding dissertations	chicago history	
Subject (Topical,Geographic)	$<c_3^n[\#]\ [[\#]c_1^n[\#]]>$			periodicals information information science
LCCN	$n_1^8[\#]$ or $n^2\text{-}n_1^6[\#]$	7523795	56-11002	75-640174
ISBN	$n_1^{10}[\#]$ or n^9x	0824019913		
Call Number	$c_1^{40}[\#]$			

Figure 4.4

University of Chicago

Access Point	Search Key	Example 1	Example 2	Example 3
Personal Name (Author,Subject)	$c_1^{253}[\#]$	Johnson, Ron 1937-# Picasso Pablo 1881-1973#	Ellis John Tracy, 1905-#	
Corporate/Conference Name (Author, Series, Subject)	$c_1^{253}[\#]$		Chicago. University. Charles R Walgreen Foundation Lectures#	American Society for Information Science#
Title	$c^3c^1c^1c^1$	EARSOP	AMECØØ	JOUOTA
Author/Title	c^3c^3	JOHEAR	ELLAME	AMEJOU
Series Title	$c_1^{253}[\#]$	Outstanding dissertations in the fine arts#	Chicago history of American Civilization	
Subject	$c_1^{253}[\#]$			Information science Periodicals# Documentation-- Periodicals#
LCCN	n_3^8	7523795	5611002	75640174
ISBN	n^{10}	0824019903		
ISSN	n^8			00028231

Figure 4.5

Library of Congress - MUMS

Access Point	Search Key	Example 1	Example 2	Example 3
Personal Name (Author, Subject)	c_1^n,c_1^m or c_1^m or c_1^5,c^1 or c_1^6	Johnson, Ron Picasso, Pabl	Ellis, J	
Corporate/Conference Name (Author, Series)	$<w_1^n>$ with Boolean coordination		Chicago university	american information science
Title	$c_1^3[\cdot]c[\cdot]c[\cdot]c$ or $<w_1^n>$ with Boolean coordination	ear,s,o,p or picasso sculpture	ame,c or american catholicism	jou,o,t,a or journal american science
Author/Title	$c_1^3[\cdot]c_1^3$	Joh,Ear	ell,ame	ame,jou
Series Title	$<w_1^n>$ with Boolean coordination	outstanding fine arts dissertations	chicago history american walgreen lectures	
Subject (Topical, Geographic)	$<w_1^n>$ with Boolean coordination		history Catholic United States	information science publications
LCCN	$[a_1^3]$ n^2-n_1^6	75-23795	56-11002	75-640174
ISBN	n_9^{10} or n^9x or $[n-]n_1^x$-n_1^y-c^1 (where x+y=8)	0824019903 or 0-8240-1990-3		
ISSN	n^4-n^4 or n^4-n^3x			0002-8231

Figure 4.6

Library of Congress - SCORPIO

Access Point	Search Key	Example 1	Example 2	Example 3
Personal Name (Author, Subject)	$c_1^{43}\#$	Johnson, Ron Picasso, Pablo	Ellis, John T	
Corporate/Conference Name (Author,Series,Subject)	$c_1^{43}\#$		Chicago Catholic Church in	American Society for information
Title	$c_1^{43}\#$	Early sculpture of pic	American cath	journal of the am
Series Title	$c_1^{43}\#$	Outstanding dissertations	Chicago history of am	
Subject (Topical, Geographic)	$c_1^{43}\#$			Information scie
LCCN	$n^2\text{-}n_1^6$	75-23795	56-11002	75-640174
Call Number	$c_1^2 n_1^{41}\#$	NB553	BX1406	Z1007

Figure 4.7

National Library of Medicine

Access Point	Search Key	Example 1	Example 2	Example 3
Personal Name (Author, Subject)	w,w[#]	Johnson, Ron: Picasso, P:	Ellis, John T:	
Corporate/Conference Name	$<w_1^n>$ with Boolean coordination		Chicago. University. Charles: or Foundation and Walgreen	American Society for Inf: Society and Information and science
Title	c^3[/c][/c][/c] or $<w_1^n>$ with Boolean coordination	EAR/S/O/P Sculpture and Picasso	AME/C	JOU/O/T/A Journal and information and science
Series Title	c_1^{39}[#]	Outstanding diss:	Chicago history of:	
Subject	$<w_1^n>/<w_1^n>$ (MeSH)		Catholic church in:	Information science and periodicals
LCCN	$n^2-n_1^6$	75-23795	56-11002	75-640174
ISBN	$n-n_1^x-n_1^y-n^1$ (BN) (where x+y=8)	0-8240-1990-3(BN)		
ISSN	n^4-n^4 (IS)			0002-8231(IS)
Call Number	c_1^{39}[#]	NB 553:	BX 1406:	Z 1007 A51

Figure 4.8

Northwestern University

Access Point	Search Key	Example 1	Example 2	Example 3
Personal Name (Author)	$< c_1^{67}[\#]>$	Johnson r	ellis j	
Corporate/Conference Name (Author, Series)	$<c_1^{67}[\#]>$ or $<c_1^i> \not\emptyset = <c_1^j[\#]>$ (where $i+j \leq 65$)		chicago university charles r walgreen =chi[illegible]o	american society
Title	$<c_1^{67}[\#]>$	early sculpture of pic	american cath	journal of the american society for info
Author/Title	$<c_1^i> \not\emptyset + <c_1^j[\#]>$ (where $i+j \leq 65$) or $<c_1^i> \not\emptyset = <c_1^j> \not\emptyset + <c_1^k[\#]>$ (where $i+j+k \leq 63$)		ellis, john +american	
Series Title	$<c_1^{67}[\#]>$	outstanding diss	chicago history of am	
ISSN	$n^4-n^4[\#]$			0002-8231

Figure 4.9

OCLC, Inc.

Access Point	Search Key	Example 1	Example 2	Example 3
Personal Name (Author)	$c_1^4,[c_1^3],[c]$	JOHN,RON,Ø	ELLI,JOH,T	
Corporate/Conference Name (author)	$=c_1^4,[c_1^3],[c]$			=INFO,SCI,Ø 1/
Title	$c_1^3,[c_1^2],[c_1^2],[c]$	EAR,SC,OF,P	AME,CA,,	JOU,OF,TH,A AME,DO,A,Q 2/
Author/Title	$[c_1^4]$, c_1^4	JOHN,EARL	ELLI,AMER	INFO,JOUR 1/
Series Title	$c_1^3,[c_1^2],[c_1^2],[c]$	OUT,DI,IN,T	CHI,HI,OF,A	
LCCN	$[c_1^3]n_1^2-n_1^6$	75-23795	56-11002	75-640174
ISBN	n^{10}	0824019903		
ISSN	n^4-n^4			0002-8231
CODEN	c_5^6			AISJB6

1/ "American Society for" is stop-listed.

2/ Search also allowed on Preceding Entry (780).

Figure 4.10

Washington Library Network

Access Point	Search Key	Example 1	Example 2	Example 3
Personal Name (Author, Subject)	$<(w_1^{30})[\#]>$	johnson ron picasso, pablo	ellis john t#	
Corporate/Conference Name (Author,Series,Subject)	$<(w_1^{30-x})[\#]_1^x> x\leq 8$ (Keyword)		chicago--university--charles# chicago walgreen lectures	american society for info# american society information
Title	$<(w_1^{30-x})[\#]_1^x> x\leq 8$ (Keyword)	picasso 1901	american catholicism	journal american information
Author/Title	Boolean coordination		ellis john# .and. american catholicism	
Series Title	$<w_1^{30}[\#]>$	outstanding diss#	chicago hist#	
Subject (Topical)	$<w_1^{30}[\#]>$(Subfield)			periodicals--info# information science--period#
LCCN	$[a_1^3]n^2[-]n_1^6$	75-23795	5611002	75640174
ISBN	n^{10} or n_3^9#	0824019903		
ISSN	n^4-n^4 or n_3^8#			0002-8231

Figure 4.11

SEARCH ARGUMENT MATRIX

	BALLOTS	CHICAGO	LC:HUMS	LC:SCORPIO	NLM	NORTHWESTERN	OCLC	WLN
PERSONAL NAME	$([c_1^n[\emptyset]])_1^m\ c_3^n[\emptyset]$ or $([w[\emptyset]])_1^m\ c_3^n[\emptyset]$, $([c_1^n[\emptyset]])_1^m$	$c_1^{253}[\emptyset]$	c_1^n, c_1^m or c_1^m or c_1^5, c^1 or c_1^6	$c_1^{43}\emptyset$	$w,w[\emptyset]$	DIRECT ORDER: $\langle c_1^{67}[\emptyset]\rangle$	$c_1^4, [c_1^3], [c]$	$\langle (w_1^{30})[\emptyset]\rangle$
CORPORATE/ CONFERENCE NAME	$\langle (w[\emptyset])_1^n\rangle$	$c_1^{253}[\emptyset]$	$\langle w_1^n\rangle$ with Boolean Coordination	$c_1^{43}\emptyset$	$\langle w_1^n\rangle$	DIRECT ORDER: $\langle c_1^{67}[\emptyset]\rangle$ ROTATED ORDER: $\langle c_1^i\rangle\#\langle c_1^j[\emptyset]\rangle$ (where i+j≤65)	$-c_1^4, [c_1^3], [c]$	$\langle (w_1^{30-x})[\emptyset]_1^x\rangle$ x≤8 (Keyword)
TITLE	$\langle (w[\emptyset])_1^n\rangle$	$c^3c^1c^1c^1$	$c_1^3[,]c[,]c[,]c$ or $\langle w_1^n\rangle$	$c_1^{43}\emptyset$	$c^3[/c][/c][/c]$ or $\langle w_1^n\rangle$	$\langle c_1^{67}[\emptyset]\rangle$	$c_1^3, [c_1^2], [c_1^2], [c]$	$\langle (w_1^{30-x})[\emptyset]_1^x\rangle$ x≤8 (Keyword)
AUTHOR/TITLE	Boolean Coordination	c^3c^3	$c_1^3[,]c_1^3$			DIRECT ORDER: $\langle c_1^i\rangle\#\langle c_1^j[\emptyset]\rangle$ (i+j≤65) ROTATED ORDER: $\langle c_1^i\rangle\#\langle c_1^j\#c_1^k[\emptyset]\rangle$ (where i+j+k≤63)	$[c_1^4], c_1^4$	Boolean Coordination
SERIES TITLE	$\langle (w[\emptyset])_1^n\rangle$	$c_1^{253}[\emptyset]$	$\langle w_1^n\rangle$	$c_1^{43}\emptyset$	$c_1^{39}[\emptyset]$	$\langle c_1^{67}[\emptyset]\rangle$	$c_1^3, [c_1^2], [c_1^2], [c]$	$\langle (w_1^{30})[\emptyset]\rangle$
SUBJECT	$\langle c_3^n[\emptyset][[\emptyset]c_1^n[\emptyset]]\rangle$	$c_1^{253}[\emptyset]$	$\langle w_1^n\rangle$	$c_1^{43}\emptyset$	$\langle w_1^n\rangle/\langle w_1^n\rangle$ (MeSH)			$\langle (w_1^{30})[\emptyset]\rangle$ (Subfield)
LCCN	$n_1^8[\emptyset]$ or $n_2-n_1^6[\emptyset]$	n_3^8	$[a_1^3]n^2-n_1^6$	$n^2-n_1^6$	$n^2-n_1^6$		$[c_1^3]n_1^2-n_1^6$	$[a_1^3]n^2[-]n_1^6$
ISBN	$n_1^{10}[\emptyset]$ or n^9x	n^{10}	n_9^{10} or n^9x or $[n-]n_1^x-n_1^y-c^1$ (where x+y=8)		$n-n_1^x-n_1^y-n^1$ (BN) (where x+y=8)		n^{10}	n^{10} or $n_3^9\emptyset$
ISSN		n^8	n^4-n^4 or n^4-n^3x		n^4-n^4 (IS)	$n^4-n^4[\emptyset]$	n^4-n^4	n^4-n^4 or $n_3^7\emptyset$
CODEN							c_5^6	
CALL NO.	$c_1^{40}[\emptyset]$			$c_1^2n_1^{41}\emptyset$	$c_1^{39}\emptyset$			

Figure 4.12

5.0 Status/Error Message Format

During the processing of search queries it is possible that errors might occur, and status/error messages must be transmitted between systems. This study investigated the types of status/error messages anticipated and their relationship to the NCLIS/NBS Protocol. The protocol has certain specific message types for conveying information about the status of the systems and to notify a system that an error or exception has occurred or that an established resource limit has been exceeded.

5.1 Status Messages

The Status Query Message (Message Type 5) and its associated Status Query Response Message (Message Type 6) allow one host to determine the current status of another host regarding a session. The Status Query Response Message includes an indication of the last message the host has received, the account number and process identification of this session, and resource use information, as well as status code. The Status Field contains a registered number indicating the status of the current session and/or last message.

It is anticipated that after a search query has been sent from one host to another, and a long delay is encountered, the requesting host would probably send a Status Query Message to the processing host to see if the search is still being processed. It is therefore felt that certain Status Codes should be registered indicating the following conditions with respect to a requested search:

Search is awaiting start - This response indicates that the request for search has been received at the target facility but for some reason has not yet actually begun execution. This might be because of low priority, necessity to mount special data bases, or some other reason.

Search is being conducted - This status response indicates that the requested search is being processed at the target facility.

Search has been completed and reply is awaiting transmission - This status response indicates that the requested search has been successfully concluded or that a logical break point has occurred in that search, and that the response is awaiting transmission. It is usually anticipated that the response to a bibliographic search will automatically be transmitted to the requesting host, but in some cases the response might be delayed.

5.2 Error Messages

The Exception Message (Message Type 11) is sent by a host whenever it detects an unusual or exceptional condition as a result of, or as a by-product of, processing a message from another host. Exception Messages can be sent because of invalid header field values, resource limits exceeded, or invalid data in the data field. Status Field values will be registered indicating the type of error/exception. It is anticipated that status codes will be registered for the following conditions associated with Search Queries:

Ill-formed command - In the event that the target facility determines that the syntax of a command it has received is improperly formatted, this status response will be transmitted in reply.

Unrecognized command - In the event that the target facility computer detects a properly formatted message but does not recognize the command code as a registered command, this status response is transmitted in reply.

Unsupported command - In the event that the target facility computer system receives a properly formatted command which exists within the national registry but which calls for an action that is not provided by that particular target facility, this status response will be transmitted in reply.

Permanent system limit exceeded - This status response indicates that some resource at the target facility which is required to satisfy the search is not available. An example of this might be an insufficiently large system buffer. This condition will terminate processing of the search and notify the other system.

5.3 Suspend Messages

The Suspend Message (Message Type 12) is sent by a host to inform another host that it has temporarily ceased processing a message pending further information from the requesting host. A possible cause of a suspend message when processing a search command might be that the established resource limit has been exceeded. The NCLIS/NBS Protocol allows the setting of a resource limit on a session or message. This resource limit on a search command prevents excessive processing and its associated costs. It is therefore possible that the following suspend status conditions should be registered:

Message Resource (Price) Limit Exceeded - The processing system has exceeded the resource limit (in price) established when the search was sent. The suspend command notifies the other system that in order to continue processing, the

message resource limit will have to be increased. The message resource limit can be increased by transmitting a Resume message wit' an increased Message Resource Limit Field.

Temporary system limit exceeded - This status response indicates that a nonpermanent system resource required for successful completion of the search is not available in the target system. One example of such a resource might be available scratch disk storage. The suspend command notifies the originating host of this condition and allows him to determine if the system should wait until sufficient resources are available and resume processing, or terminate processing the requested search.

5.4 Informatory Messages

Some status or exception type responses will be normal informatory messages sent by the bibliographic system and handled as normal responses. These include:

No record matches search criteria - This response indicates that the target facility data base contains no record that satisfies the search criteria.

6.0 Search Query Responses

The issue of response formats to bibliographic search queries is very complex. After this study was begun, it was decided that response formats should be investigated in another separate study because of their complexity.

6.1 Response Contents

This additional study should first determine what data are contained in different responses.

Number of records - This response indicates the number of records which match the search criteria. This is a very simple response that usually leads to further search refinement and record display.

Abbreviated record(s) - This response contains only an abbreviated record or records which match the search criteria. It will have to be determined which data elements of a bibliographic record should be contained in an abbreviated record.

Complete record(s) - This record contains the complete bibliographic record which matches the search criteria.

6.2 Response Formats

The format for the bibliographic response will also have to be determined. Most on-line systems now transmit messages in "screen format" to their terminals. These formats would have to be modified for computer-to-computer transmission, probably into a format based on the ANSI Z39.2-1979 standard. (American National Standard for Bibliographic Information Interchange on Magnetic Tape)

Probably the most complex issue concerning responses is the hierarchical nature of search/response interactions. On most systems a subsequent search query can be based upon the response received from an earlier search query. For example, if a search query yields several records, most systems display a few records (usually abbreviated) and allow the user to select the desired record; if the search query yields a large number of records, a more complex set of procedures allows the user to browse and select the desired record by a series of commands and responses. The comparison of the various response formats, command/response techniques, and search refinement should be performed in a future study. This study should include:

- multiple record response formats
- sorting techniques

- record selection techniques
- hierarchical searching techniques
- search argument combinations

7.0 Recommendations and Conclusions

A brief analysis of the search query formats shows that there are major differences between systems in the formats for personal, corporate, and conference name, title, author/title, and subject, and close similarities in formats for LCCN, ISBN, and ISSN.

Personal name search queries are usually in inverted order (surname first) and use a truncated string, keyword, or specially formed search key (4,3,1-key).

Corporate/conference name queries are based on a truncated string, keyword, or specially formed key (4,3,1-key).

Title queries are based on either truncated string, keyword, or specially formed keys (3,1,1,1-key 3,2,2,1-key).

Author/title queries are based either on a boolean combination or personal, coporate, or conference name query with the title query, or on a specially formed key (3,3-key or 4,4-key).

Subject queries are based on a truncated string, a keyword technique, or a special combination of field-subfield keys indicating hierarchical nature of subject headings.

LC Card Number queries are based on the eight digits with the optional three character prefix and hyphen.

International Standard Book Number queries are based on the ten digits with optional hyphens.

International Standard Serial Number queries are based on the eight digits with optional hyphens.

Only one system allows search on CODEN.

Call number queries are based on truncated classification and book number.

7.1 Standard Queries

Because this report focuses on search argument formats for computer-to-computer communications, some of the nuances or conveniences of search arguments that are provided for easy human input are not applicable. It is therefore recommended that the following search argument formats be made standard in initial implementations of computer-to-computer communications.

LC Card Number - $a^{3}n^{8}$

This format is an exact match of the 001 field in the MARC format.

Example: 75-23795 becomes ␣␣␣75023795.

ISBN - n^{10} or $n^{9}x$

This format is an exact match of the 020 field in the MARC format.

Example: 0-8240-1990-3 becomes 0824019903.

ISSN - n^{8} or $n^{7}x$

Example: 0002-8231 becomes 00028231.

For the other search arguments, it is recommended that representatives of the major bibliographic systems meet to decide if standard formats can be implemented and what these formats should be. It is felt that several search argument formats may have to be standardized and registered. These would include:

Truncated string - the access point as it is appears in the record with truncation permitted after an agreed upon number of characters.

Keyword - combinations of one or more keywords from the access point.

Special Key - a specially formed key made of a select number of characters of various words in the access point.

Probably not all systems would be able to implement all standard queries initially, but systems with similar search techniques would be expected to modify current queries to implement the standard for computer-to-computer communications.

7.2 Response Formats

It is recommended that a separate study be undertaken to investigate the various response formats and search session interaction characteristics as described previously in Section 6.0. This study should identify the different responses generated by systems, the method by which searches are refined, and desired items found. The study should determine whether the commonalities between the various responses could be standardized for computer-to-computer communications.

7.3 Conclusions

This study of bibliographic search queries and their associated status/error messages is just one step in the evolution of a comprehensive nationwide bibliographic system. The major bibliographic utilities must continue to work together to resolve the differences in the search argument formats, to implement and refine the NCLIS/NBS Protocol, and to develop the necessary telecommunications links to permit the transfer of messages among the various systems. Through this cooperative activity, users of one bibliographic utility will have access to the additional data and services of other systems.

OTHER NETWORK PLANNING PAPERS

No. 1 Butler, Brett. A Nationwide Location Data Base and Service. (Washington, Library of Congress, 1978). 66 p. ED168 516

No. 2 Dataflow Systems. A Glossary for Library Networking. (Washington, Library of Congress, 1978). 34 p. ED168 516

No. 3 Buchinski, Edwin J. Initial Considerations for a Nationwide Data Base. (Washington, Library of Congress, 1978). 56 p. ED168517

No. 4 Network Technical Architecture Group. Message Delivery System for the National Library and Information Service Network: General Requirements. (Washington, Library of Congress, 1978). 35 p. ED168 518

Available by request from the Customer Services Section, Cataloging Distribution Service, Library of Congress, Navy Yard Annex, Bldg. 159, Washington, D.C. 20541.